Catalytic and biophysical investigation of rhodium hydroformylase

Catalytic and biophysical investigation of rhodium hydroformylase

Hasan T. Imam[a,c], Amanda G. Jarvis*[b], Veronica Celorrio[d], Irshad Baig[a], Christopher C. R. Allen[e], Andrew C. Marr*[c] and Paul C. J. Kamer*[f]

[a]School of Chemistry, University of St Andrews, EaStCHEM, St Andrews, Fife, UK
[b]School of Chemistry, University of Edinburgh, EaStCHEM, Edinburgh, UK
[c]School of Chemistry and Chemical Engineering, Queen's University Belfast, Belfast, UK
[d]Diamond Light Source Ltd, Diamond House, Harwell Campus, Didcot, UK
[e]School of Biological Sciences, Queen's University Belfast, Belfast, UK
[f]Bioinspired Homo- & Heterogeneous Catalysis, Leibniz Institute for Catalysis, Rostock, Germany

**Corresponding authors: Paul.Kamer@catalysis.de, a.marr@qub.ac.uk, amanda.jarvis@ed.ac.uk*

Key words: Artificial metalloenzymes, Rhodium, Hydroformylation, Protein characterization

Abstract

Rh-containing artificial metalloenzymes based on two mutants of Sterol Carrier Protein_2L (SCP_2L) have been shown to act as hydroformylases, exhibiting significant activity and unexpectedly high selectivity in the hydroformylation of a range of alkenes. Here we report modifications of the catalyst performance by site directed mutagenesis, and studies of the biophysical properties of these catalysts. Catalysts were prepared based on single methionine mutants of SCP_2L for hydroformylation studies. Multinuclear (^{1}H, & ^{13}C), multi-dimensional (2D) solution NMR spectroscopy, EXAFS and XANES were employed to probe the structure. Biophysical studies using 2D [^{1}H ^{13}C] HSQC NMR spectroscopy of ^{13}C-methyl methionine labeled catalysts were used to investigate changes in protein conformation. Hydroformylation studies employing the methionine mutants as catalysts revealed the significant effects of mutation on hydroformylation activity. Among the methionine mutant catalysts M1A of V83C and A100C, and M112A of V83C exhibited significantly higher activity than their parent proteins with improved selectivity. A metal binding role for methionine was suggested by EXAFS and XANES data on selenomethionine variants.

Introduction

Chemical production methods strongly depend on tailor-made catalysts. Enzymes, nature's catalysts, provide superior performance in many catalytic reactions. Nevertheless, many industrial production processes are reliant on chemical conversions such as CO and alkene insertions, for which no natural enzymes are available. In the area of synthetic chemistry, control over the activities and selectivities of reactions is key to increasing sustainability, and learning how to design more efficient catalysts is one of the major challenges in green catalysis. Artificial metalloenzymes (ArM) utilize bio-macromolecules as ligands to bind metal or metal-containing prosthetic groups and perform unprecedented chemical catalysis under environmentally benign conditions. [1–10] Design of ArMs that promote higher selectivity and activity is challenging. Strategies to achieve this have included, supramolecular assembly,[11,12] creating metal binding sites *in vivo* and reconstituted with a metal ion,[13,14] metal substitution,[15–17] incorporation of unnatural metal binding amino acids,[10,18–20] computational or *in silico*[21] and covalent attachment including phosphonate ester and thioester linkage,[22,23] alkylation or maleimide conjugation.[24,25] One of the major advantages of ArMs is the great scope for tuning the catalyst activity and selectivity by genetic optimization i.e. site-directed mutagenesis on the protein scaffold, the so-called second coordination sphere.[26]

Artificial metalloenzymes containing phosphines and their derivatives have received growing interest due to the wide versatility of these ligands in homogeneous catalysis, and in particular their use in bulk chemical processes such as hydroformylation and hydrogenation. Pioneering work was reported by Wilson and Whitesides, [1] who incorporated biotinylated diphosphine Rh complexes into the protein avidin and generated a hybrid catalyst for the hydrogenation of acetamidoacrylic acid with modest enantiomeric excess of up to 44% for the *S* isomer. Hydrogenation of itaconic acid by a hybrid catalyst of biotinylated Pyrphos-Rh complex with avidin was found to be active, giving up to 99% conversion, however it gave modest enantiomeric excess of up to 44% for the *R* isomer.[27] ArMs constructed from biotinylated Rh-diphosphine complexes with a protein moiety of avidin, neutravidin or streptavidin, among them streptavidin-biotin-Rh-diphosphine catalysts, were found to be active in the hydrogenation of acetamidoacrylic acid affording quantitative conversion with 92% enantiomeric excess of *R* isomer. Mutation on streptavidin at S112G enhanced the enantiomeric excess of *R* isomer up to 96%.[28] Direct evolution of the Rh-diphosphine-avidin hybrid mutant on S112G, yielded double mutants N49V/S112G and N49/V giving 35%, 54% and 65% enantiomeric excess of the *R* isomer, respectively, in the hydrogenation of acetamido acrylic ester.[29] Diphosphine-Rh complex [(1,5-cyclooctadiene){bis(2-diphenylphosphinoethyl)succinamido}rhodium(I)]perchlorate was conjugated with keyhole limpet hemocyanin (KLH) or bovine serum albumin (BSA) to raise antibodies for binding the Rh-phosphine complexes. The antibody-Rh-phosphine hybrid catalysts were applied in the hydrogenation of 2-acetamidoacrylic acid. The Antibody IG8-Rh catalyst provided an enantiomeric excess of 98% towards the *S* isomer with 23% yield. However, the catalyst was only suitable for the small substrate size of haptan.[30]

Hydroformylation (or the oxo process) is a large-scale catalytic reaction. Olefins are converted into aldehydes with the addition of H and a formyl group (CHO) across the double bond. The reaction has 100% atom economy and is catalyzed by transition metal catalysts (usually Co or Rh). Aldehydes are precursors for many other reactions including oxidation, hydrogenation and

reductive amination, and have applications in the lubricants, plasticizers and polyamides industries. In seminal work Wilkinson and co-workers[31] concluded that rhodium phosphine catalysts enhanced the activity and selectivity of hydroformylation catalysis under mild conditions. Following this, numerous studies have focused on the exploitation of the electro-stereochemical properties of phosphorus containing ligands, including studies of - bite angles,[32,33] mono- and bi-dentate phosphoramidites and phosphinites,[34] phosphites,[35–37] phosphine-phosphites,[38] amphiphilic phosphine ligands,[39] bidentate diphosphines,[40,41] and phosphacyclic diphosphines.[42] Much progress has been made, but challenges remain in product separation and metal leaching from immobilized catalysts. Distillation is a commonly used method for separating the product, this is suitable for several low-boiling products, but is limited by the thermal stability of the catalyst. Catalyst immobilization has been reported for Rh hydroformylation employing a wide variety of media including ionic liquids,[43,44] solid supported silica or silicate matrices,[45–47] porous organic polymers,[48] micelles, mesoporous silicas,[49] propylene carbonate,[50] super critical CO_2,[51] and biopolymers.[52] Aqueous biphasic rhodium catalyzed hydroformylation of propene using a water-soluble Rh-TPPTS catalyst is one of the most efficient approaches applied industrially to addressing product separation and catalyst recycling. For higher alkenes, however, the process becomes economically unviable because of the low solubility of the alkene resulting in low reaction rates. ArMs are a recent addition to the spectrum of hydroformylation catalysts, with successful protein-complex combinations being termed *hydroformylases*.

Artificial metalloenzymes of human serum albumin (HSA), papain and egg albumin were developed for the hydroformylation of olefins by mixing proteins with the Rh-metal complex [Rh(acac)$(CO)_2$]. At optimized reaction conditions of 70 bar (CO/H_2, 1:1), 60°C and 24 h, almost quantitative conversion towards aldehyde was observed for styrene and 1-octene, however, linear to branched selectivity was found to be 5/95 and 47/53, respectively. The low linear to branch ratio suggests that free Rh was present under the reaction conditions.[53,54] Apo carbonic anhydrase, prepared by dialysis against Zn chelator 2,6-pyridinedicarboxylate, substituted with Rh, was found to be active in the hydroformylation of styrene at room temperature for 20 h under 20 bar of synthesis gas (H_2/CO, 1:1). Modification of histidine and/or mutation at the histidines resulted in catalyst variants with different protein to Rh ratios. The variant DEPC-H4/10R+H17F-[Rh] gave 89/11 linear to branched product. With the same variant of DEPC-H4/10R+H17F-[Zn]-[Rh], keeping the Zn in the active site and Rh on the protein surface favored the branched aldehyde with linear to branch regioselectivity of 22/78, suggesting catalytically active metal in the active metal binding site of the protein favors the linear product.[55] However, direct metalation of apo carbonic anhydrase with various Rh precursors resulted in Rh-CA complexes which did not exhibit any catalytic activity towards hydrogenation and hydroformylation of a range of internal, terminal alkenes and vinyl amines. The authors termed the Rh-CA complexes as discrete organometallic protein complexes.[17]

We previously reported an artificial hydroformylase[56] based on a lipid binding protein called Sterol Carrier Protein_2L (SCP_2L) (1IKT[57], Fig 1), which showed good activity (TONs up to 400 at 48h) and unexpectedly high regioselectivity (up to 80/20 linear/branched aldehyde) in the hydroformylation of long-chain alkenes under biphasic conditions. An EXAFS study suggested single phosphine coordination to Rh and additional sulfur coordination from a methionine near the Rh center. SCP_2L has four methionines in positions M1, M80, M105 and M112. We hypothesized that by identifying which methionine was near the rhodium we could build up a more

detailed understanding of the hydroformylase active site, and thus optimize the activity and selectivity of the hydroformylase.

As the equivalent of the chemical catalyst's ligand, the protein plays a very significant role in the catalyst performance. Unlike many conventional ligands, proteins are very sensitive to chemical conditions (temperature, pressure, organic solvents and stirring) thus the ArM catalyst stability is largely dependent on the protein stability (within experimental conditions) for optimal performance, hence, it is extremely important to consider protein stability when designing an ArM catalyst. Loss of protein structure can be observed as a loss of selectivity due to the leaching of metal, or a loss of activity due to protein precipitation. Herein we describe a study of the catalyst structure and stability. Single methionine mutants were prepared and further hydroformylation studies were performed. Biophysical studies were carried out by monitoring the methionine chemical shifts during catalyst preparation using multinuclear (^{1}H & ^{13}C) & multi-dimensional (2D) solution NMR spectroscopy. EXAFS and XANES were performed on SeMet variants.

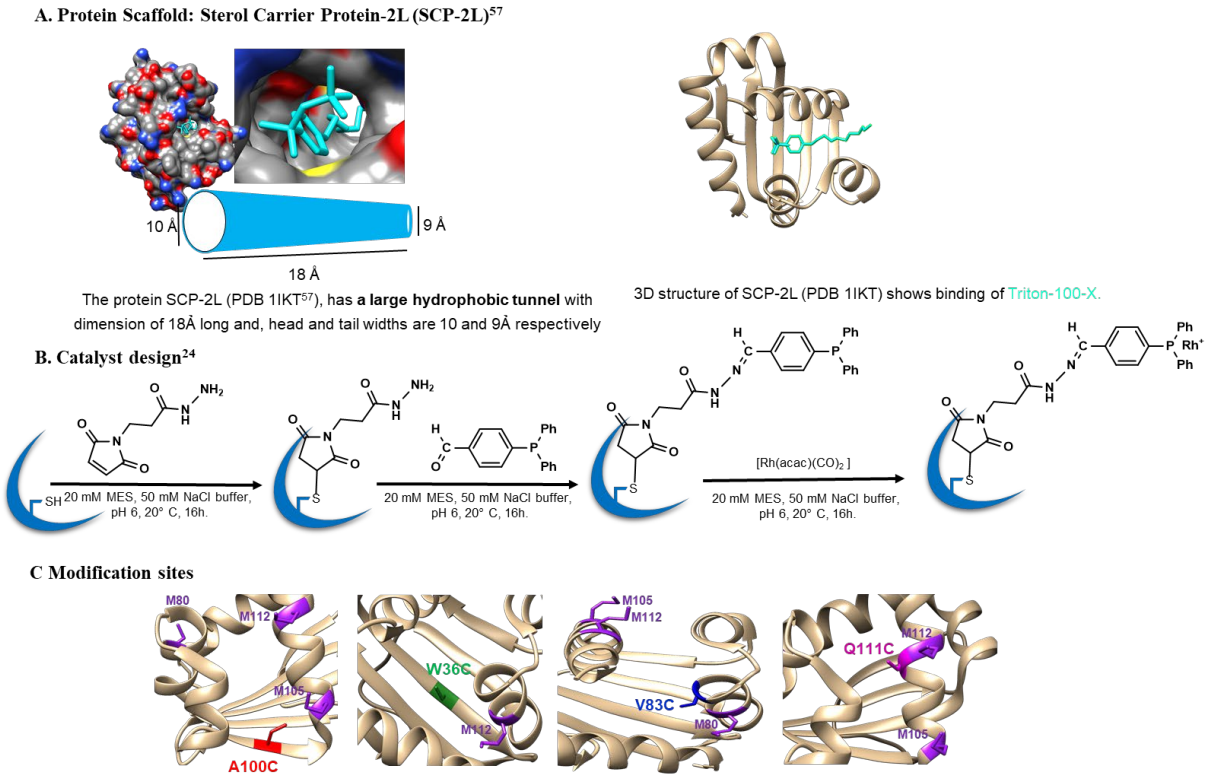

Fig 1 Protein scaffold and catalyst design, A) Structure of SCP-2L (PDB: 1IKT) showing the hydrophobic tunnel, substrate binding B) Schematic presentation catalyst design. C) four modification sites with methionine.

Results and discussion

The natural mammalian protein Sterol Carrier Protein SCP_2L, has a large hydrophobic tunnel of 18 Å long with a diameter of 10 Å, allowing it to bind hydrophobic molecules. In order to probe if modification within the tunnel could improve the selectivity, cysteine residues were introduced by site directed mutagenesis at two further residues Q111C and W36C, and the corresponding proteins were expressed in reasonable yields. SCP_2L-Q111C-**1** (glutamine at position 111 in SCP_2L was replaced by cysteine) was prepared by covalent modification using the previously described robust maleimide-phosphine conjugation method[24,56] (Fig 1B and S.I. S1.2), and mass spectra of the protein were collected (Figs S1 and S2). W36C was successfully obtained but unfortunately introduction of the organometallic cofactor was unsuccessful; the obtained maximum conversion was 20% (Fig S3) and all attempts to optimize failed. It is hypothesized that this is due to the position of W36 in the center of the tunnel and hindered access for the substrates to the cysteine, or collapse of the protein fold and thus the tunnel. Addition of $[Rh(acac)(CO)_2]$ to SCP_2L-Q111C-**1** provided SCP_2L-Q111C-**1**-P-Rh, as confirmed by mass spectrometry (Fig S2). The protein was treated with half an equivalent of Rh with respect to protein concentration, leaving available Rh binding sites to avoid non-specific binding, and washed three times with degassed buffer employing a 10 kDa MWCO centrifugal filter to remove free Rh. The activity of SCP_2L-Q111C-**1**-P-Rh was tested in the biphasic hydroformylation of 1-octene and found to afford 252 TON after 45h, and a linear/ branched aldehyde ratio of 72/28 (Table S2). These results are in line with those previously obtained for SCP_2L-V83C-1-P-Rh ((TON 75, l/b 78/22 at 48h) and SCP_2L-A100C-1-P-Rh (TON 408, l/b 79/21 at 48 h).[56]

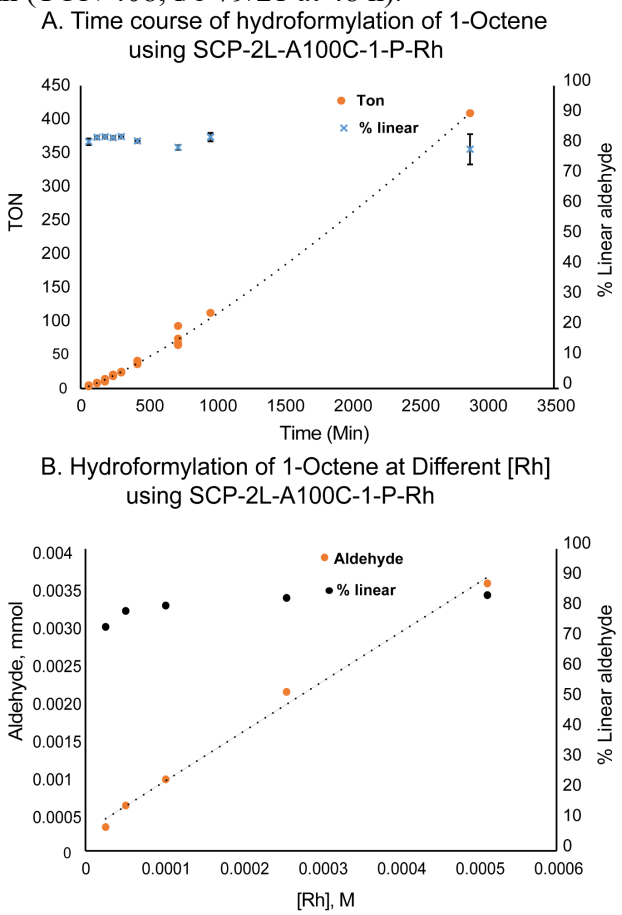

Fig 2 Catalytic performance of SCP-2L-A100C-1-P-Rh. A) Time dependent hydroformylation of 1-octene. Reaction conditions: 80 bar syngas (1:1), 35°C, 630 rpm. Samples: catalyst: 500 µl, substrate: 450 µl (1-octene), internal standard: 50 µl (9% v/v n-heptane, diphenyl ether, 1%, from 10% stock). Time points after 720 minutes were conducted using a different batch of catalyst. B) Aldehyde product formed at different [Rh]

In order to gain a greater understanding of the promising catalyst SCP_2L-A100C-1-P-Rh, we investigated the activity in the hydroformylation of 1-octene as a function of time and the effect of ArM concentration. The results from the time course experiments are summarized in Fig 2A. It was not possible to sample the reaction over time, due to the need to conduct the experiments in an autoclave under pressure. Therefore, each hydroformylation experiment was carried out individually for different time periods using fresh catalyst from the same batch. The selectivity towards the linear aldehyde was constant across the first 5 h (between 82-83%), with a slight decrease by 48 h (79 ±4 %,).
In some hydroformylase reactions protein precipitation was observed after a 48 h reaction time. Conditions were optimized to ensuring good protein stability and solubility, as these are key factors in ArM performance. The metalloenzymes remained soluble in water and in the buffer (20 mM MES, 50 mM NaCl, pH 6.0), and no products/species were observed due to protein degradation or hydrolysis of -N=C- bonds, as evidenced from mass spectra (Fig S2 and Fig S12). Under the experimental conditions (80 bar syn gas, 35°C, 630 rpm), the metalloenzyme SCP-2L-100C-1-P-Rh was active for 48h, however, we have seen minor drops in selectivity (82% at 5h to 79% at 48h, Fig 2A). The relatively high selectivities obtained (82-79%, at wide range of time 5-48h) suggest that the ArMs are reasonable stable under the experimental conditions.

The TON increased in an essentially linear fashion (R^2 = 0.99, on a linear trendline). This is indicative of pseudo zeroth order in substrate and steady state, and is thus suggestive of a stable catalyst. Closer examination reveals a slight increase in rate over time (Fig S15). This could be due to an incubation period, either due to an initial slow rate whilst the protein binds the substrate (lag phase), or due to an incubation time for hydride formation,[39] with higher incubation time accelerating the product formation. At the end of the reaction time (48 h, 2880 min) the catalyst was still active. The effects of changing the ArM catalyst concentration are represented by Fig 2B. As the concentration of catalyst increased so did the concentration of aldehyde produced. Doubling the concentration of catalyst roughly doubled the concentration of aldehyde, suggesting an approximately first order dependence on catalyst concentration under these conditions.

In our previous studies the interpretation of EXAFS data suggested that Rh binds to one of the methionines of SCP_2L derived hydroformylases.[56] To further explore the role of methionine in hydroformylase activity, alanine scanning mutagenesis was carried out on both SCP_2L-V83C-1-P-Rh and SCP_2L-A100C-1-P-Rh.[56] In each case one methionine was mutated at a time, and the subsequent modified mutants (SCP_2L-V83C-MnA-1-P-Rh and SCP_2L-A100C-MnA-1-P-Rh, where n is the original position of the methionine of 1, 80, 105 or 112) tested for the hydroformylation of 1-octene. The results are summarized in Fig 3 and Table S5. Unfortunately, SCP_2L-V83C-M80A was not obtained in high enough yields to progress to catalytic studies. The catalyst SCP-2L-V83C-1-P-Rh (Fig 3A) exhibited a TON of 6.2 at 16 h with regioselectivity of 72/28 towards the linear and branched aldehyde. The three mutants SCP_2L-V83C-M1A-1-P-Rh, SCP_2L-V83C-M105A-1-P-Rh and SCP_2L-V83C-M112A-1-P-Rh all gave TONs which were considerably higher than the parent. The M1A variant gave the best result (TON 58±22 at 16 h). The regioselectivities for M1A and M112A were similar to V83C (72/28 and 75/25 vs 76/24 linear to branched aldehyde), with M105A exhibiting a lower regioselectivity of 69/31 linear to branched

product. Small amounts of Rh leaching cannot be ruled out when lower selectivities are observed, as free Rh is less selective, therefore this may point to a lower stability for the M105A mutant.

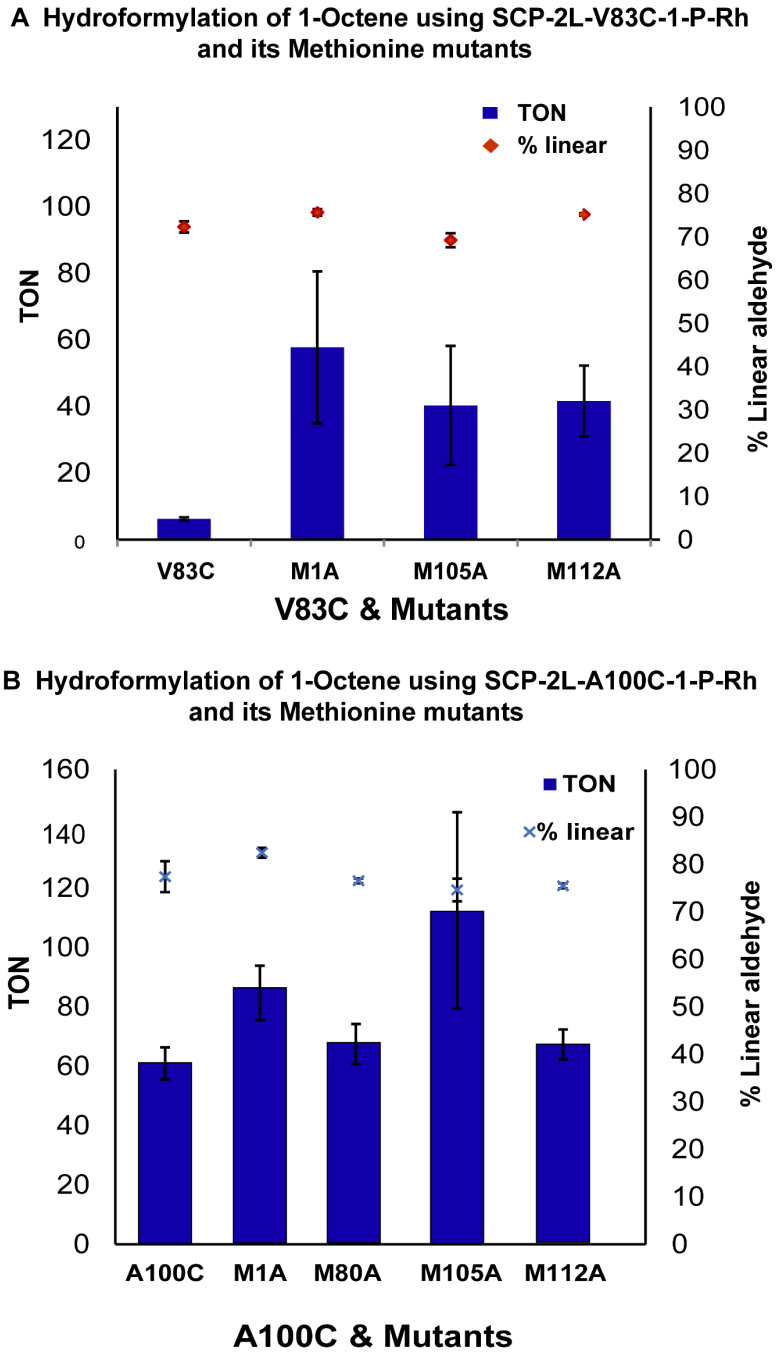

Fig 3 Hydroformylation of 1-Octene by SCP-2L variant ArM. (A) SCP-2L-V83C-1-P-Rh and its methionine mutant's variant, activities and selectivities shown in blue square and red diamond respectively. (B) SCP-2L-A100C-1-P-Rh and its methionine variant, activities and selectivities shown in blue square and cross, respectively Reaction Conditions: 80 bar syngas (1:1), 35°C, 630 rpm, 16 hours. Catalyst: 500 µl, (Rh concentration V83C & Mutants, V83C 38.7, M1A 37.0, M105A 20.4 and M112A 24.5µM, Rh concentration A100C& Mutants A100C 20.0, M1A 16.2, M80A 18.5, M105A 18.2 and M112A 14.0 µM), substrate: 450 µl (1-Octene), Internal standard: 50 µl (9% v/v n-heptane, diphenyl ether, 1%, from 10% stock). Activity and selectivity were measured by GC, 3-4 replicates. Errors are reported as standard mean of error.

The results for the SCP_2L-A100C derived proteins are not as dramatic (Fig 3B). Two of the mutants performing within experimental error of the parent. However SCP_2L-A100C-M1A and M105A both exhibited significantly higher TONs at 16 h compared to SCP_2L-A100C. M105A gave higher TONs, but the results must be viewed with caution, as there was a wide variation of TONs measured, as indicated from the statistical error (±33), and this ArM had the lowest average regioselectivity of 74/26 toward the linear and branched product. This could indicate problems in the stability of the mutant. In contrast SCP_2L-A100C-M1A-1-P-Rh performed consistently and gave the highest selectivity of 81/19, linear to branched aldehyde. This represents the most successful catalyst in this part of the study.

For both the catalyst variants, the replacement of a methionine led to an increase in the TON and slight changes in selectivity. For SCP_2L-A100C-M1nA-1-P-Rh mutants, as each methionine was replaced, including, presumably the methionine that was suggested[56] to bind to Rh, the catalytic performance improved, or was consistent with the parent. This would appear to rule out an activating role for methionine as a promotor in the catalytic cycle. The nature of the catalyst appears to be highly complex, with changes made across the protein having an influence on the catalytic activity. In this way ArM catalysts are more akin to enzymes than homogeneous catalysts, and therefore the study of ArMs must be the study of the whole polypeptide, and not just the region immediately around the metal. A further postulate was that M105 was anchoring Rh in SCP_2L-A100C-1-P-Rh, leading to a relatively stable resting state. The removal of methionine M105 and replacement with alanine, which lacks good donors, may therefore destabilize the ArM, but also enable it to perform at a higher rate, if somewhat erratically.

In order to shed more light on the role of methionine the carbon and hydrogen environments were studied by NMR spectroscopy. Biomolecular solution NMR spectroscopy is an extremely useful technique for monitoring the solution behavior of biomolecules. 2D [^{1}H ^{13}C] HSQC solution NMR spectroscopy of ^{13}C-methyl-methionine labelled protein is a powerful method that has been previously used to investigate metal-methionine binding and protein conformational studies.[58–61] Proteins labelled with ^{13}C-methyl-methionine were expressed and purified in order to monitor how the methionine chemical shift changed as the protein was covalently modified. 2D [^{1}H ^{13}C] HSQC solution NMR spectroscopy was employed to monitor the methyl chemical shift as the catalyst was prepared.

The protein has four methionines (M1, M80, M105 and M112), so it was important to identify and assign the methionine residues. For the methionine assignment a 2D [^{1}H ^{13}C] HSQC NMR spectrum of the ^{13}C-methyl-methionine labelled SCP-2L-V83C was recorded at 303K (Fig 4A). With the aid of spectra of single methionine mutants (Fig 4B-D), it was possible to assign the methionines unambiguously. In the spectrum of ^{13}C-Methyl-Met-V83C (Fig 4A), there are four cross peaks- marked as 1,2,3 and 4- observed from the four methyl resonances. Replacement of M80 with alanine (Fig 4B) results in the disappearance of peak 1, suggesting this cross peak is for M80.

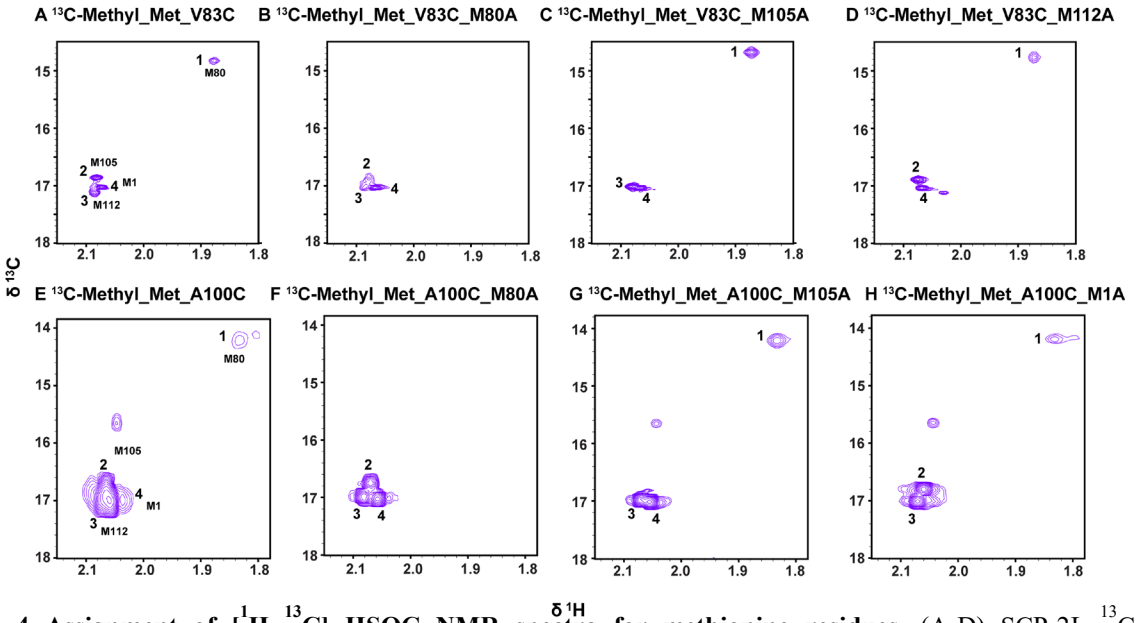

Fig 4 Assignment of [^{1}H ^{13}C] HSQC NMR spectra for methionine residues. (A-D) SCP-2L_^{13}C-Methyl_Met_V83C and its methionine mutants and (E-H) SCP-2L_^{13}C-Methyl_Met_A100C and its methionine mutants.

Similarly, mutation M105A (Fig 4C) and M112A (Fig 4D) resulted in the absence of the respective cross peaks thus facilitating the assignment. As M80, M105 and M112 resonances were assigned unambiguously, the remaining peak 4, was assigned to M1. Similar experiments were performed with ^{13}C-Methyl_Met_A100C and its methionine mutants (Fig 4 E-H), mutated A100C_M1A confirm the M1 peak assignment. The methionine peak positions for the two variants were very similar, M80 had a very small chemical shift difference (of 0.6 ppm) in ^{13}C NMR. Each methionine mutation resulted in small chemical shift changes of the other methionine resonances, these are summarized in Table 1.

Table 1 NMR chemical shift of ^{13}C and ^{1}H of ^{13}C-methyl-methionine of SCP-2L-V83C and SCP-2L-A100C and their methionine mutatants

Entry	M1	M80	M105	M112
	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm
SCP-2L-V83C	17.02/2.06	14.74/1.86	16.86/2.06	17.05/2.07
SCP-2L-V83C-M1A		-		
SCP-2L-V83C-M80A	17.03/2.06		16.88/2.07	17.02/2.08
SCP-2L-V83C-M105A	17.03/2.06	14.68/1.87		17.02/2.07
SCP-2L-V83C-M112A	17.03/2.07	14.76/1.87	16.88/2.07	
SCP-2L-A100C	16.99/2.03	14.21/1.83	16.75/2.06	17.07/2.07
SCP-2L-A100C-M1A		14.19/1.83	16.80/2.06	17.01/2.07
SCP-2L-A100C-M80A	17.02/2.05		16.74/2.07	16.98/2.08
SCP-2L-A100C-M105A	16.80/2.06	14.20/1.83	-	16.99/2.06
SCP-2L-A100C-M1112A				

The effects of methionine mutation on protein folding were investigated by 1D ^{1}H NMR spectroscopy (Fig S11). The aromatic or NH regions (6-11 ppm) of the spectra of V83C and A100C and their mutant proteins look very similar. Resonances in the aromatic region (6-11 ppm) are particularly well dispersed, suggesting a well folded protein. In the aromatic regions some

resonances became sharper and higher in intensity upon mutation, suggesting these residues were buried.

Catalyst preparation was monitored by recording 2D [^{1}H ^{13}C] HSQC NMR spectra of ^{13}C-methyl-methionine labelled SCP-2L variants of V83C and A100C. The catalyst preparation was a three steps process (Fig 1B), addition of maleimide (1) to form the protein-maleimide conjugate, then addition of a phosphine ligand (P), which attaches to the protein-maleimide ensemble and finally coordination of a rhodium moiety (Rh).

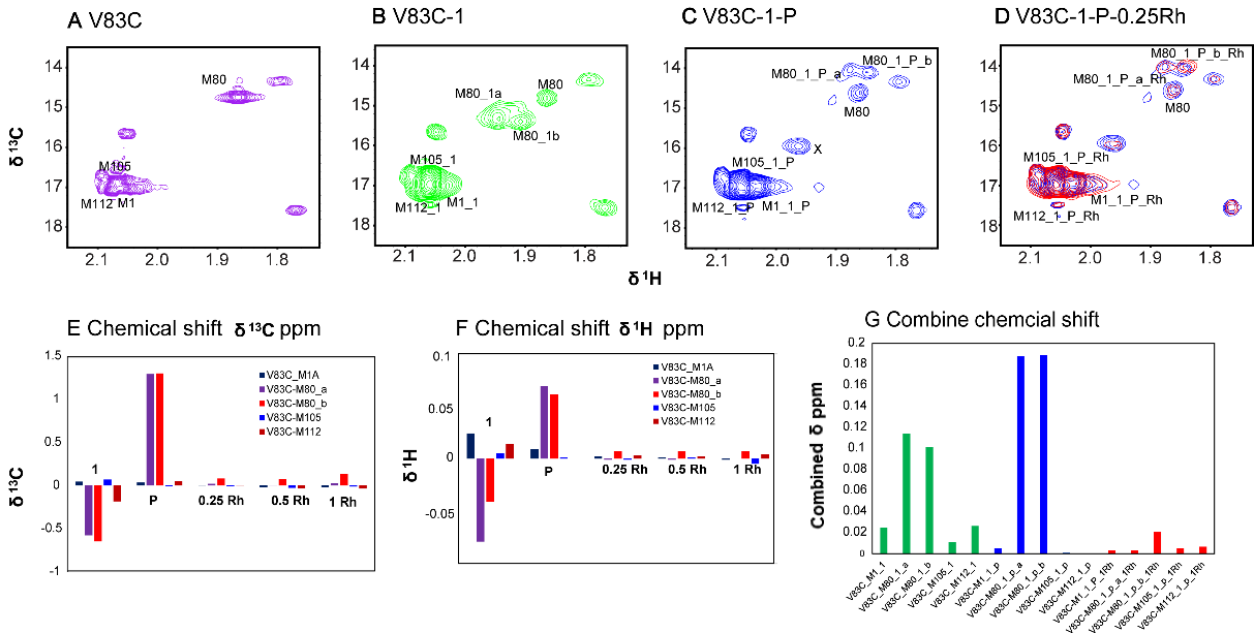

Fig 5 Monitoring the catalyst design by [^{1}H ^{13}C] HSQC NMR spectroscopy. (A) SCP-2L_^{13}C-Methyl_Met_V83C (B) Maleimide (1) modified SCP-2L_^{13}C-Methyl_Met_V83C, (C) Phosphine (P) conjugated maleimide (1) modified SCP-2L_^{13}C-Methyl_Met_V83C. (D) Addition of 0.25 equivalent of [Rh(acac)$(CO)_2$] (Rh) to phosphine-maleimide- SCP-2L_^{13}C-Methyl_Met_V83C. Changes in chemical shifts (ppm) of methionine methyl group due to covalent modifications (E) Chemical shift change in δ^{13}C dimension (F) Chemical shift change in δ ^{1}H dimension (G) combined chemical shifts δ(combine). CSP calculated based on the equation δppm = $((\Delta^{1}H)^{2}+0.142(\Delta^{13}C)^{2})^{0.5}$

Table 2 NMR chemical shift of ^{13}C and ^{1}H of ^{13}C-methyl-methionine of SCP-2L-V83C due to chemical modification during catalyst design

Entry	M1	M80	M105	M112
	$^{13}C/^{1}H$ ppm	$^{13}C/^{1}H$ ppm	$^{13}C/^{1}H$ ppm	$^{13}C/^{1}H$ ppm
SCP-2L-V83C	17.02/2.06	14.74/1.86	16.86/2.06	17.05/2.07
SCP-2L-V83C-1	16.98/2.04	15.32/1.94[a]	16.79/2.06	17.23/2.06
SCP-2L-V83C-1		15.39/1.90[b]		
SCP-2L-V83C-1-P	16.94/2.03	14.03/1.87[a]	16.80/2.06	17.18/2.06
SCP-2L-V83C-1-P		14.08/1.84[b]		
SCP-2L-V83C-1-P-0.25Rh	16.94/2.02	14.01/1.87[a]	16.80/2.06	17.18/2.06
SCP-2L-V83C-1-P-0.25Rh		14.01/1.84[b]		
SCP-2L-V83C-1-P-0.50Rh	16.96/2.02	14.03/1.87[a]	16.82/2.06	17.21/2.06
SCP-2L-V83C-1-P-0.50Rh		14.02/1.84[b]		
SCP-2L-V83C-1-P-1Rh	16.96/2.03	14.01/1.87[a]	16.81/2.06	17.22/2.06
SCP-2L-V83C-1-P-1Rh		13.96/1.84[b]		

a and b are the two peaks assign for M80 due to modification

Methyl chemical shifts due to the chemical modification of SCP-2L-V83C are shown in Fig 5 with the chemical shifts of methionine cross peaks as the protein was modified. Chemical shifts are listed in Table 2. Apart from resonances assigned to the methionines, three other peaks were apparent, these peaks may be due to methyl resonance from other methyl containing amino acids, perhaps from alanine or isoleucine, considering that their methyl chemical shifts usually appear close to the methyl chemical shift of methionine (Table 1, examples of typical alanine and isoleucine methyl chemical shifts: A 18.96/1.36 ppm, I 13.40/0.68 and 17.53/0.78 ppm in $^{13}C/^{1}H$ NMR[62]). Maleimide modification (Fig 5B) resulted in a small chemical shift of M80, however, for the other three methionine residues M1, M105 and M112 no significant chemical shifts were observed. In the region of M80, two new peaks appeared, the two resonances may be due to the diastereomers formed on maleimide binding. These two peaks are labeled as M80_1a and M80_1b. In Fig 5E and F chemical shifts changes in the ^{1}H and ^{13}C NMR spectra due to maleimide, phosphine modification and Rh addition are displayed. Maleimide modification resulted in small ^{13}C chemical shift differences for M80_1a and M80_1b. Phosphine modification resulted in small shifts of the peaks M80_1a and M80_1b in the opposite direction labelled as M80_1_P_a and M80_1_P_b. Addition of one equivalent of $[Rh(acac)(CO)_2]$ resulted in very small chemical shifts of the ^{13}C NMR resonance M80_1_P_b, while for ^{1}H NMR the change was negligible. The small chemical shift changes observed are not conclusive, however, a chemical shift change of 1.3 ppm in ^{13}C NMR observed due to phosphine modification could suggest that the protein experiences a small local conformation change around the M80. The crystal structure of SCP-2L[57] indicates that the M80 site is populated with both hydrophobic/aromatic and charged amino acid residues, so the maleimide-phosphine modification could induce both electrostatic and hydrophobic interactions.

The spectra and methionine chemical shifts for SCP-2L-A100C upon modification are displayed in (Fig S13) and summarized in Table S6. Maleimide and phosphine conjugation (Fig S13) has little or no effect on the methionine resonances. In addition, no significant changes in the methionine resonances were observed due to Rh addition. A close inspection of the mass spectrum (S12G) suggests (~5%) Rh loading in this catalyst, and this may explain the small shifts. However, in reality the Rh content could be higher, as previously observed for Photoactive Yellow Protein (PYP).[23]

Protein folding upon catalyst formation was investigated by 1D ^{1}H NMR. Analysis of the amide/ aromatic region of the 1D ^{1}H NMR data suggested protein conformation changes as we prepare the catalyst (Fig 6).

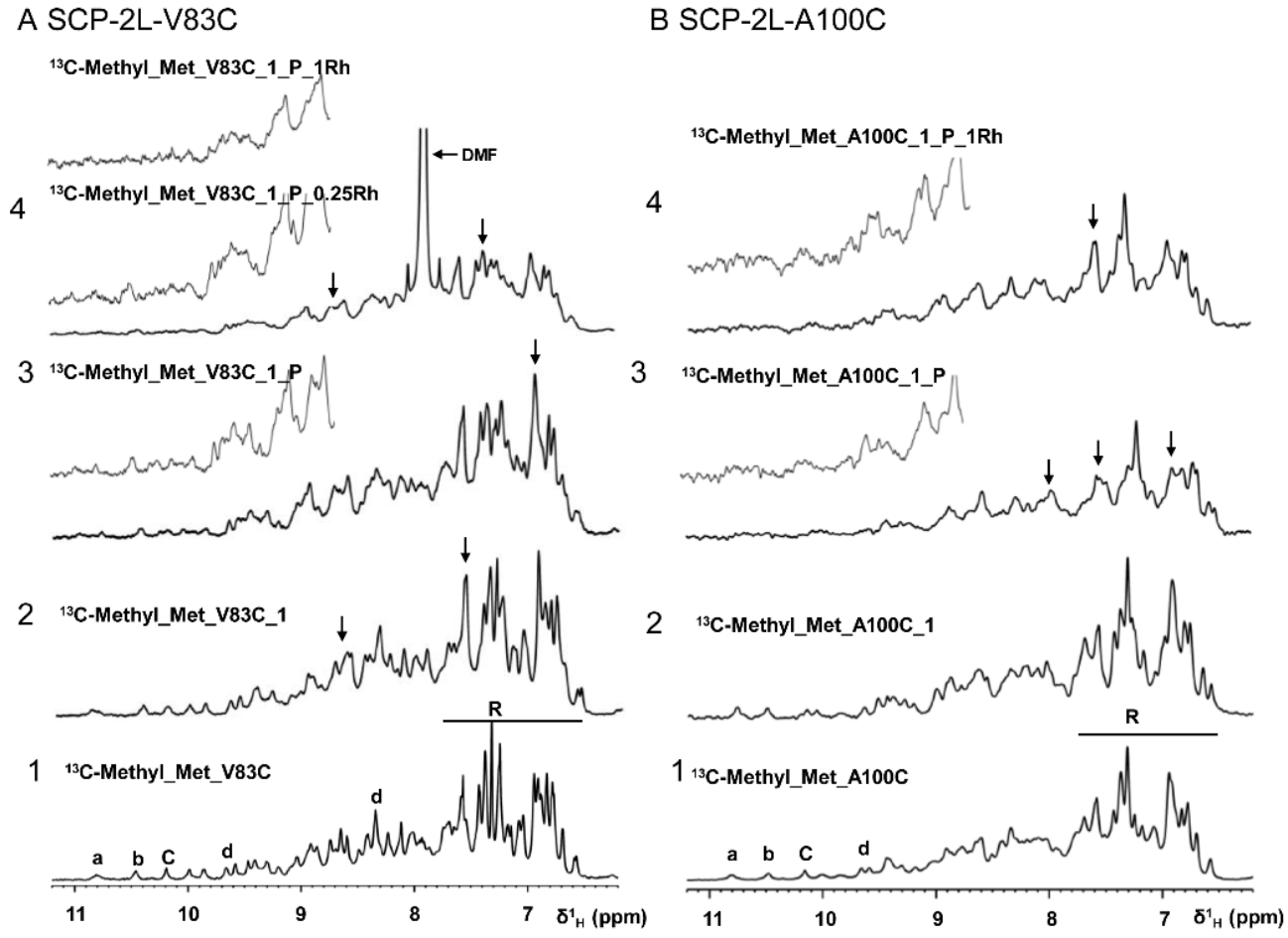

Fig 6 Monitoring the protein folding by 1D 1H NMR spectroscopy as prepare the catalyst. (A) SCP-2L_^{13}C-Methyl_Met_V83C (B) (A) SCP-2L_^{13}C-Methyl_Met_A00C.

Unmodified protein exhibited good dispersion and a high intensity peak (Fig 6A1 and 6B1) suggesting a well folded protein. As the protein was modified with maleimide some peak broadening was observed (Fig 6A2 and 6B2 marked with arrow in the region R). Peaks assigned as a, b, c and d were still sharp and visible, suggesting the proteins remain mostly folded after maleimide modification. Drastic changes in peak dispersion and intensity were observed in the phosphine modified protein (Fig 6A3 and 6B3). Peaks a, b, c and d were broadened or disappeared. Peak broadening was also observed for protons belonging to the region of R, which are the aromatic protons, comprising the hydrophobic core, this peak broadening is suggestive of the disruption of hydrophobicity. For A100C, the effect is more severe and significant peak broadening was observed. The lower peak intensity for A100C could also be a function of the low

concentration, however, the magnified spectrum indicates peaks disappearance or broadening. With Rh addition, both V83C and A100C peaks in regions a, b, c and d broaden or disappear suggesting partial change in protein folding or conformation.

It was hoped biophysical studies by 2D [^{1}H ^{13}C] HSQC would reveal features of methionine behavior upon protein modification and these could assist understanding of the trends in hydroformylation activity. The M1A methionine mutants SCP_2L-V83C-M1A-1-P-Rh and SCP_2L-A100C-M1A-1-P-Rh were particularly successful catalysts compared to their parent proteins. However, our biophysical studies by NMR spectroscopy did not provide any direct evidence of M1-Rh binding or M1-modifier interactions from the chemical shift changes. In the crystal structure of SCP-2L,[57] M1 is found in the flexible region of the protein and did not appear in the structure. Serine 8 (S8) is in the helix of the protein, within 6.7Å of the modification site C83, as S8 and M1 are separated by 6 amino acids and due to M1's flexible nature, M1 could be in the proximity of the modification site. Therefore, under hydroformylation conditions, a direct role for methionine M1, in which the sulfur hinders hydroformylation, cannot be ruled out. We were not able to perform catalytic reactions on an M80 mutant due to low protein yield, this could be because M80 mutation results in low stability, as NMR studies suggest that this site is sensitive to chemical modification. For M105 and M112 no significant and reliable changes in chemical shift were observed due to modification nor Rh addition. In conclusion no direct evidence for methionine to Rhodium binding in V83C variants could be obtained from the NMR study. Similarly protein modification of A100C variants had very little or no impact on the chemical shift of methionine residues, and very small chemical shift changes were observed for M105 in ^{13}C NMR.

To determine if methionine M105 could be the methionine responsible for metal binding in A100C, we performed EXAFS and XANES experiments on SCP_2L(SeMet)-A100C-M105A-1-P-Rh. In this variant the methionines in SCP_2L-A100C-M105A-1-P-Rh were replaced by selenomethiones, making them more visible in the XAFS spectra. The EXAFS and XANES spectra then gave information on the environment around Rh. The study is described in depth in the S.I. section S1.3.5. The results show a change in the XANES and EXAFS compared to SCP_2L(SeMet)-A100C-1-P-Rh. However, in common with the previous study,[56] the data for SCP_2L(SeMet)-A100C-M105A-1-P-Rh suggests the presence of a selenium atom close to the Rh center. It is therefore unlikely that the sulfur of M105 is coordinated to Rh in SCP_2L-A100C-1-P-Rh. Therefore we propose that the increase in catalytic activity observed for SCP-2L-A100C-M105A-1-P-Rh is not due to the loss of a S-Rh interaction in the resting state of the catalyst. This provides more evidence that enhancements in rate can occur due to changes in amino acids remote from the metal centre. We postulate that enhancement of catalytic efficiency in ArMs is subtle, and can be affected by conformational changes upon substitution of single amino acids.

Conclusions

Using our published ArMs as a starting point,[56] SCP-2L-A100C-1-P-Rh and SCP-2L-V83C-1-P-Rh, mutants have been expressed and purified in order to probe structure function relationships in SCP-2L derived hydroformylases. In particular the role of methionine has been investigated, and the effects of protein modification on methionine have been probed using NMR techniques. By careful modification and optimization it has been found that artificial hydroformylases can be prepared that are active for up to 48 hours, such as SCP-2L-A100C-1-P-Rh. The precise amino

acid sequence has a profound effect on the stability and catalytic efficiency of the ArM, as it does in enzymes. The positioning of the metal binding site within the protein scaffold has a significant impact on the catalytic performance. This may be because the different amino acid residue environment around Rh affects substrate binding. Engineered protein scaffolds with single methionine mutants enabled the enhancement of the catalytic activity. In both the SCP-2L-V83C and SCP-2L-A100C variants, enhancements in the activity and selectivity were observed for M1A mutation. For SCP-2L-A100C M105A mutation led to considerable increases in activity, but a slight loss in selectivity. EXAFS studies on SCP_2L(SeMet)-A100C-M105A-1-P-Rh detected Rh-Se coordination, and suggested that M105 is not directly involved in Rh binding in SCP-2L-A100C-1-P-Rh. A complex picture is emerging from these combined studies, and we postulate that catalytic performance can be radically affected by conformational changes in the protein brought about by alterations in amino acids distant from the active site. In this way ArM catalysts are more akin to enzymes than homogeneous catalysts. These studies reveal the true complexity of ArM catalysts, as every change to the amino acid sequence, and the spacer, linker and metal site, can have effects beyond the immediate proximity of the change. The future of ArM research will therefore depend, not only on the choice of good protein scaffolds to support catalytic activity, but also on the precise design of polypeptides that embrace and enhance the desired catalytic activity.

Conflicts of Interest

Authors have no conflicts of interest

Acknowledgements

HTI, AM, PCJK would like to thank The EPSRC UK Catalysis hub for funding this project through EPSRC grant reference EP/M013219/1. UK Catalysis Hub is kindly thanked for resources and support provided via our membership of the UK Catalysis Hub Consortium and funded by EPSRC grants: EP/K014706/2, EP/K014668/1, EP/K014854/1, EP/K014714/1 and EP/M013219/1**.** AGJ thanks the EU for support through a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014-657755) and the University of Edinburgh for a Christina Miller fellowship. We thank Dr Siobhan Smith & Dr Tomas Lebl, University of St-Andrews and Dr Juraj Bella University of Edinburgh for helping with NMR experiments. We thank Dr Sally Shirran and Dr Silvia Synowsky, University of St-Andrews for helping with Mass spectrometry. We thank UK catalysis Hub and Diamond light source for the opportunity to perform EXAFS and XANES studies (*via* B18BAG funding). The authors wish to acknowledge the Diamond Light Source for provision of beamtime (SP19850). HTI thanks Mr Peter McNiece for presenting the work at the ISGC-2019.

References

1 M. E. Wilson and G. M. Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.
2 F. Rosati and G. Roelfes, *ChemCatChem*, 2010, **2**, 916–927.
3 T. Heinisch and T. R. Ward, *Curr. Opin. Chem. Biol.*, 2010, **14**, 184–199.
4 P. J. Deuss, R. Denheeten, W. Laan and P. C. J. Kamer, *Chem. Eur. J.*, 2011, **17**, 4680–4698.
5 A. Pordea, *Curr. Opin. Chem. Biol.*, 2015, **25**, 124–132.

6 O. Pàmies, M. Diéguez and J. E. Bäckvall, *Adv. Synth. Catal.*, 2015, **357**, 1567–1586.
7 T. K. Hyster and T. R. Ward, *Angew. Chem. Int. Ed.*, 2016, **55**, 7344–7357.
8 F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis and T. R. Ward, *Chem. Rev.*, 2018, **118**, 142–231.
9 M. Jeschek, S. Panke and T. R. Ward, *Trends Biotechnol.*, 2018, **36**, 60–72.
10 U. Markel, D. F. Sauer, J. Schiffels, J. Okuda and U. Schwaneberg, *Angew. Chem. Int. Ed.*, 2019, **58**, 4454–4464.
11 J. Bos, W. R. Browne, A. J. M. Driessen and G. Roelfes, *J. Am. Chem. Soc.*, 2015, **137**, 9796–9799.
12 A. Pordea, M. Creus, C. Letondor, A. Ivanova and T. R. Ward, *Inorganica Chim. Acta*, 2010, **363**, 601–604.
13 T. R. Ward, *Angew. Chem. Int. Ed.*, 2016, **55**, 14909–14911.
14 H. M. Key, P. Dydio, D. S. Clark and J. F. Hartwig, *Nature*, 2016, **534**, 534–537.
15 Q. Jing, K. Okrasa and R. J. Kazlauskas, *Chem. Eur. J.*, 2009, **15**, 1370–1376.
16 M. Ohashi, T. Koshiyama, T. Ueno, M. Yanase, H. Fujii and Y. Watanabe, *Angew. Chem. Int. Ed.*, 2003, **42**, 1005–1008.
17 H. M. Key, D. S. Clark and J. F. Hartwig, *J. Am. Chem. Soc.*, 2015, **137**, 8261–8268.
18 I. S. Hassan, A. N. Ta, M. W. Danneman, N. Semakul, M. Burns, C. H. Basch, V. N. Dippon, B. R. McNaughton and T. Rovis, *J. Am. Chem. Soc.*, 2019, **141**, 4815–4819.
19 S. Hanna, J. C. Holder and J. F. Hartwig, *Angew. Chem. Int. Ed.*, 2019, **58**, 3368–3372.
20 P. Srivastava, H. Yang, K. Ellis-Guardiola and J. C. Lewis, *Nat. Commun.*, 2015, **6**, 1–8.
21 T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker and T. R. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 10414–10419.
22 M. Basauri-Molina, C. F. Riemersma, M. A. Würdemann, H. Kleijn and R. J. M. Klein Gebbink, *Chem. Commun.*, 2015, **51**, 6792–6795.
23 W. Laan, B. K. Muñoz, R. Den Heeten and P. C. J. Kamer, *ChemBioChem*, 2010, **11**, 1236–1239.
24 P. J. Deuss, G. Popa, C. H. Botting, W. Laan and P. C. J. Kamer, *Angew. Chem. Int. Ed.*, 2010, **49**, 5315–5317.
25 H. Osseili, D. F. Sauer, K. Beckerle, M. Arlt, T. Himiyama, T. Polen, A. Onoda, U. Schwaneberg, T. Hayashi and J. Okuda, *Beilstein J. Org. Chem.*, 2016, **12**, 1314–1321.
26 A. Pordea and T. R. Ward, *Chem. Commun.*, 2008, 4239–4249.
27 C. C. Lin, C. W. Lin and A. S. C. Chan, *Tetrahedron Asymmetry*, 1999, **10**, 1887–1893.
28 J. Collot, J. Gradinaru, N. Humbert, M. Skander, A. Zocchi and T. R. Ward, *J. Am. Chem. Soc.*, 2003, **125**, 9030–9031.
29 M. T. Reetz, J. J. P. Peyralans, A. Maichele, Y. Fu and M. Maywald, *Chem. Commun.*, 2006, 4318–4320.
30 H. Yamaguchi, T. Hirano, H. Kiminami, D. Taura and A. Harada, *Org. Biomol. Chem.*, 2006, **4**, 3571–3573.
31 D. Evans, J. A. Osborn and G. Wilkinson, *J. Chem. Soc. A*, 1968, 3133–3142.
32 P. W. N. M. van Leeuwen, P. C. J. Kamer and J. N. H. Reek, *Pure Appl. Chem.*, 2007, **71**, 1443–1452.
33 L. A. Van Der Veen, M. D. K. Boele, F. R. Bregman, P. C. J. Kamer, P. W. N. M. Van Leeuwen, K. Goubitz, J. Fraanje, H. Schenk and C. Bo, *J. Am. Chem. Soc.*, 1998, **120**, 11616–11626.
34 S. C. Van der Slot, J. Duran, J. Luten, P. C. J. Kamer and P. W. N. M. Van Leeuwen,

Organometallics, 2002, **21**, 3873–3883.
35 A. van Rooy, E. N. Orij, P. C. J. Kamer and P. W. N. M. van Leeuwen, *Organometallics*, 1995, **14**, 34–43.
36 G. J. H. Buisman, M. E. Martin, E. J. Vos, A. Klootwijk, P. C. J. Kamer and P. W. N. M. van Leeuwen, *Tetrahedron: Asymmetry*, 1995, **6**, 719–738.
37 A. Van Rooy, P. C. J. Kamer, P. W. N. M. Van Leeuwen, K. Goubitz, J. Fraanje, N. Veldman and A. L. Spek, *Organometallics*, 1996, **15**, 835–847.
38 A. Van Rooy, J. N. H. De Bruijn, K. F. Roobeek, P. C. J. Kamer and P. W. N. M. Van Leeuwen, *J. Organomet. Chem.*, 1996, **507**, 69–73.
39 M. S. Goedheijt, B. E. Hanson, J. N. H. Reek, P. C. J. Kamer and P. W. N. M. Van Leeuwen, *J. Am. Chem. Soc.*, 2000, **122**, 1650–1657.
40 L. A. Van Der Veen, P. C. J. Kamer and P. W. N. M. Van Leeuwen, *Organometallics*, 1999, **18**, 4765–4777.
41 P. C. J. Kamer, P. W. N. M. van Leeuwen and J. N. H. Reek, *Acc. Chem. Res.*, 2001, **34**, 895–904.
42 A. Buhling, P. C. J. Kamer and P. W. N. M. van Leeuwen, *J. Mol. Catal. A. Chem.*, 1995, **98**, 69–80.
43 X. Jin, K. Zhao, F. Cui, F. Kong and Q. Liu, *Green Chem.*, 2013, **15**, 3236–3242.
44 T. E. Kunene, P. B. Webb and D. J. Cole-Hamilton, *Green Chem.*, 2011, **13**, 1476–1481.
45 L. A. Van Der Veen, P. C. J. Kamer and P. W. N. M. Van Leeuwen, *Angew. Chem. Int. Ed.*, 1999, **38**, 336–338.
46 N. J. Meehan, A. J. Sandee, J. N. H. Reek, P. C. J. Kamer, P. W. N. M. Van Leeuwen and M. Poliakoff, *Chem. Commun.*, 2000, 1497–1498.
47 A. J. Sandee, J. N. H. Reek, P. C. J. Kamer and P. W. N. M. Van Leeuwen, *J. Am. Chem. Soc.*, 2001, **123**, 8468–8476.
48 C. Li, L. Yan, L. Lu, K. Xiong, W. Wang, M. Jiang, J. Liu, X. Song, Z. Zhan, Z. Jiang and Y. Ding, *Green Chem.*, 2016, **18**, 2995–3005.
49 W. Zhou and D. He, *Green Chem.*, 2009, **11**, 1146–1154.
50 P. Neubert, S. Fuchs and A. Behr, *Green Chem.*, 2015, **17**, 4045–4052.
51 C. T. Estorach, A. Orejón and A. M. Masdeu-Bultó, *Green Chem.*, 2008, **10**, 545–552.
52 B. C. E. Makhubela, A. Jardine and G. S. Smith, *Green Chem.*, 2012, **14**, 338–347.
53 S. P. C. Bertucci, C. Botteghi, D. Giunta, M. Marchetti, *Adv. Synth. Catal.*, 2002, **344**, 556–562.
54 M. Marchetti, G. Mangano, S. Paganelli and C. Botteghi, *Tetrahedron Lett.*, 2000, **41**, 3717–3720.
55 Q. Jing and R. J. Kazlauskas, *ChemCatChem*, 2010, **2**, 953–957.
56 A. G. Jarvis, L. Obrecht, P. J. Deuss, W. Laan, E. K. Gibson, P. P. Wells and P. C. J. Kamer, *Angew. Chem. Int. Ed.*, 2017, **56**, 13596–13600.
57 A. M. Haapalainen, D. M. F. Van Aalten, G. Meriläinen, J. E. Jalonen, P. Pirilä, R. K. Wierenga, J. K. Hiltunen and T. Glumoff, *J. Mol. Biol.*, 2001, **313**, 1127–1138.
58 E. J. Beatty, M. C. Cox, T. A. Frenkiel, B. M. Tam, A. B. Mason, R. T. A. MacGillivray, P. J. Sadler and R. C. Woodworth, *Biochemistry*, 2002, **35**, 7635–7642.
59 H. Sun, M. C. Cox, H. Li, A. B. Mason, R. C. Woodworth and P. J. Sadler, *FEBS Lett.*, 1998, **422**, 315–320.
60 H. Sun, H. Li, A. B. Mason, R. C. Woodworth and P. J. Sadler, *J. Biol. Chem.*, 2001, **276**, 8829–8835.

61 K. Siivari, M. Zhang, A. G. Palmer and H. J. Vogel, *FEBS Lett.*, 1995, **366**, 104–108.
62 E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao and J. L. Markley, *Nucleic Acids Res.*, 2008, **36**, DOI:10.1093/nar/gkm957.